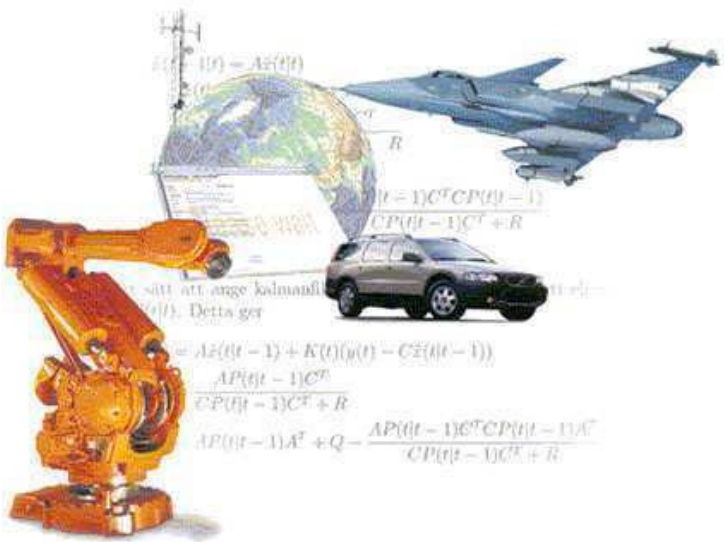
Pedestrian Group Tracking Using the GM-PHD Filter



Viktor Edman†, Maria Andersson‡, Karl Granström*, Fredrik Gustafsson*

† SAAB AB, Linköping, Sweden

‡ Division of Sensor Informatics,
Department of Sensor & EW Systems
Swedish Defence Research Agency

* Division of Automatic Control,
Department of Electrical Engineering
Linköping University, Sweden

Pedestrian group tracking

- Objectives:
  1. Track the groups of pedestrians
  2. Represent the shape of the groups
  3. Estimate the number of people in each group

Pedestrian group tracking

- Objectives:
  1. Track the groups of pedestrians
  2. Represent the shape of the groups
  3. Estimate the number of people in each group
- Tested on publicly available video data:
  “PETS 2012 dataset S1: Person count and density estimation”.

Frame 50 Frame 100 Frame 150

Different approaches to group tracking

Approaches for group tracking can roughly be divided as follows;

1. Group tracking without individual tracks.
   - Treat the group as a single object, do not track single individuals.
   - Conceptually similar to extended target tracking.

Different approaches to group tracking

Approaches for group tracking can roughly be divided as follows;

1. Group tracking without individual tracks.
   - Treat the group as a single object, do not track single individuals.
   - Conceptually similar to extended target tracking.
2. Individual target tracking supplemented by group tracking.
   - Track each individual, form groups of the estimates.

Different approaches to group tracking

Approaches for group tracking can roughly be divided as follows;

1. Group tracking without individual tracks.
   - Treat the group as a single object, do not track single individuals.
   - Conceptually similar to extended target tracking.
2. Individual target tracking supplemented by group tracking.
   - Track each individual, form groups of the estimates.

Proposed approach to group tracking

Proposed approach to group tracking;

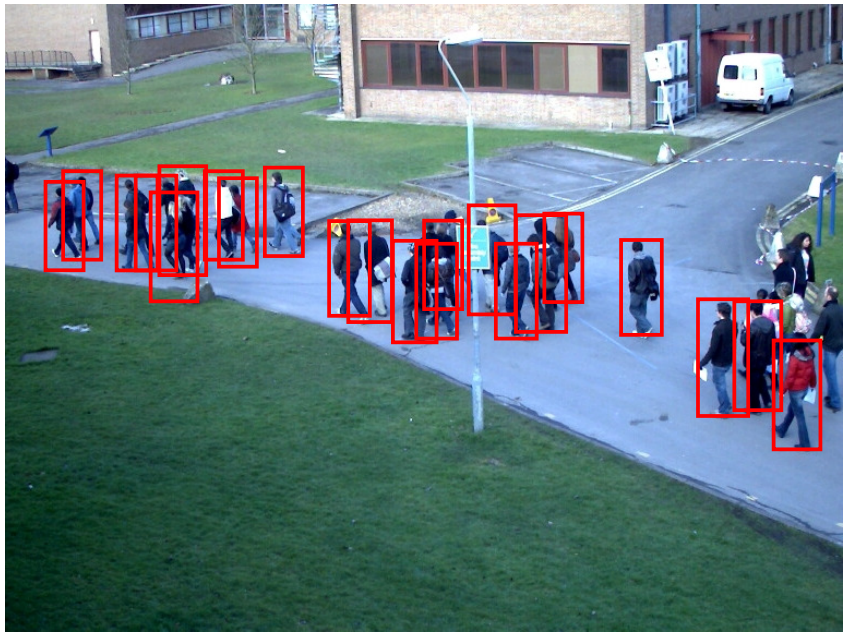
I In each image, detect pedestrians

II Transform image detections to ground positions

III Use GM-PHD filter to track the pedestrians

IV Cluster the GM-PHD filter output to form groups

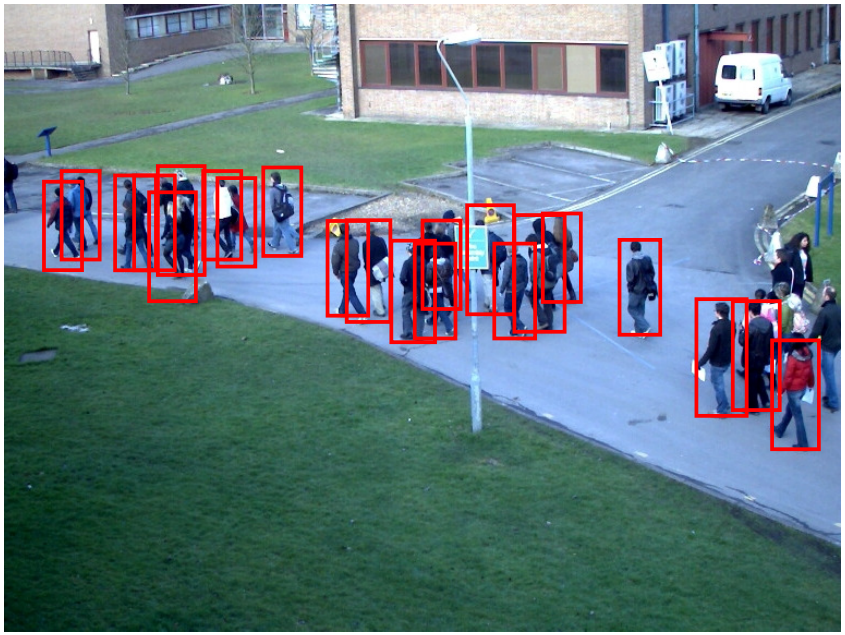
V Threshold PHD-intensity for each group to represent shape

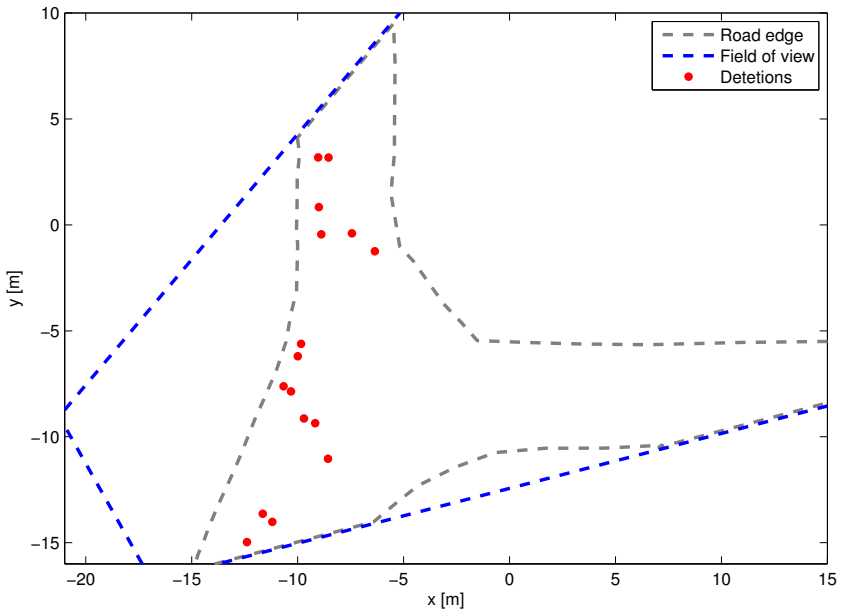


- Publicly available detector by Dollár *et al*, refs in paper
- Low measurement noise
- Missed detections due to occlusions. $p_D = 0.70$
- Few false detections (< 1 per image), i.e. low clutter

II Transform image detections to ground positions

- Center point of bottom edge projected onto ground plane.
- Requires:
  1. Calibrated camera – available online with images
  2. Known camera pose – available online with images
  3. Known ground elevation map – here world is assumed flat





- Gaussian mixture PHD filter

$$v_{k|k}(x) = \sum_{i=1}^{J_{k|k}} w_{k|k}^{(i)} \mathcal{N}\left(x; m_{k|k}^{(i)}, P_{k|k}^{(i)}\right)$$

- State vector is position and velocity: $x = [\mathrm{p}^T \ \mathrm{v}^T]^T$

- Gaussian mixture PHD filter

$$v_{k|k}(x) = \sum_{i=1}^{J_{k|k}} w_{k|k}^{(i)} \mathcal{N}\left(x; m_{k|k}^{(i)}, P_{k|k}^{(i)}\right)$$

- State vector is position and velocity: $x = [\mathrm{p}^T \ \mathrm{v}^T]^T$
- Constant velocity motion model
- Detections are linear measurements of position p_k
- New targets appear where FoV intersect the road

- Number of targets in area A and velocity region V given by

$$\int_{A,V} v_{k|k}(x)\mathrm{d}x$$

- Number of targets in area A and velocity region V given by

$$\int_{A,V} v_{k|k}(x)\mathrm{d}x \approx \sum_{i:\mathcal{N}\left(x;m_{k|k}^{(i)},P_{k|k}^{(i)}\right)\in A,V} w_{k|k}^{(i)}$$

III Use GM-PHD filter to track the pedestrians

- Number of targets in area A and velocity region V given by

$$\int_{A,V} v_{k|k}(x)\mathrm{d}x \approx \sum_{i:\mathcal{N}\left(x;m_{k|k}^{(i)},P_{k|k}^{(i)}\right)\in A,V} w_{k|k}^{(i)}$$

- Pedestrian estimates:

$$\hat{x}_{k|k}^{(i)} = m_{k|k}^{(i)} \qquad i \ : \ w_{k|k}^{(i)} > \tau$$

- Standard choice $\tau = 0.5$.

- For the extracted pedestrian estimates $\hat{x}_{k|k}^{(i)}$
- Two pedestrian estimates i and j are connected if
  1. $\left\| \hat{\mathrm{p}}_{k|k}^{(i)} - \hat{\mathrm{p}}_{k|k}^{(j)} \right\|_2 < 2$ m and
  2. $\left\| \hat{\mathrm{v}}_{k|k}^{(i)} - \hat{\mathrm{v}}_{k|k}^{(j)} \right\|_2 < 1$ m/s

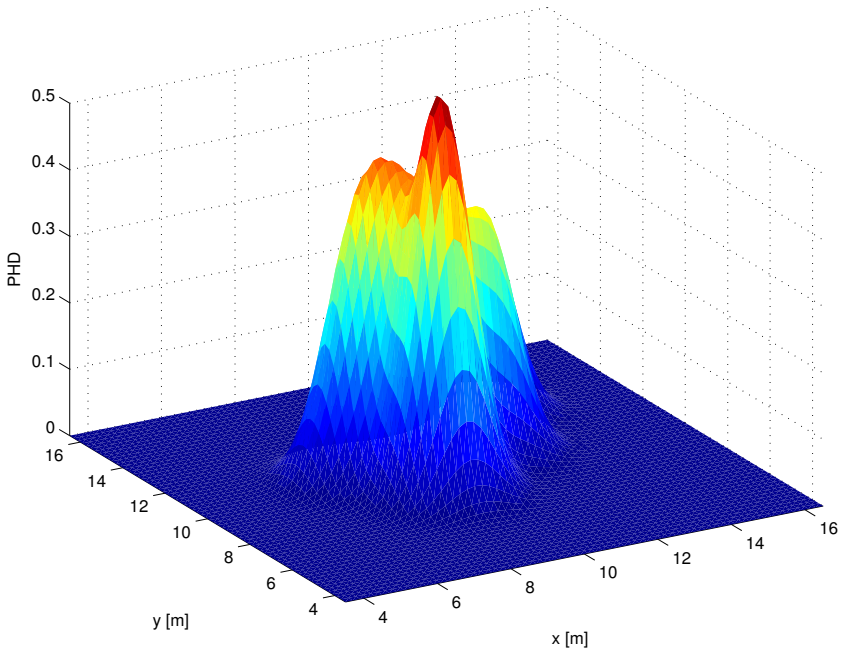
- For the extracted pedestrian estimates $\hat{x}_{k|k}^{(i)}$
- Two pedestrian estimates i and j are connected if
  1. $\left\| \hat{\mathrm{p}}_{k|k}^{(i)} - \hat{\mathrm{p}}_{k|k}^{(j)} \right\|_2 < 2$ m and
  2. $\left\| \hat{\mathrm{v}}_{k|k}^{(i)} - \hat{\mathrm{v}}_{k|k}^{(j)} \right\|_2 < 1$ m/s
- Two pedestrian estimates k and ℓ belong to the same group g if
  1. there is a set of $N \geq 2$ components indexed $i_1, i_2, \ldots, i_N$,
  2. where $i_1 = k$ and $i_N = \ell$,
  3. and where i_s and i_{s+1} are connected for $s = 1, \ldots, N-1$.

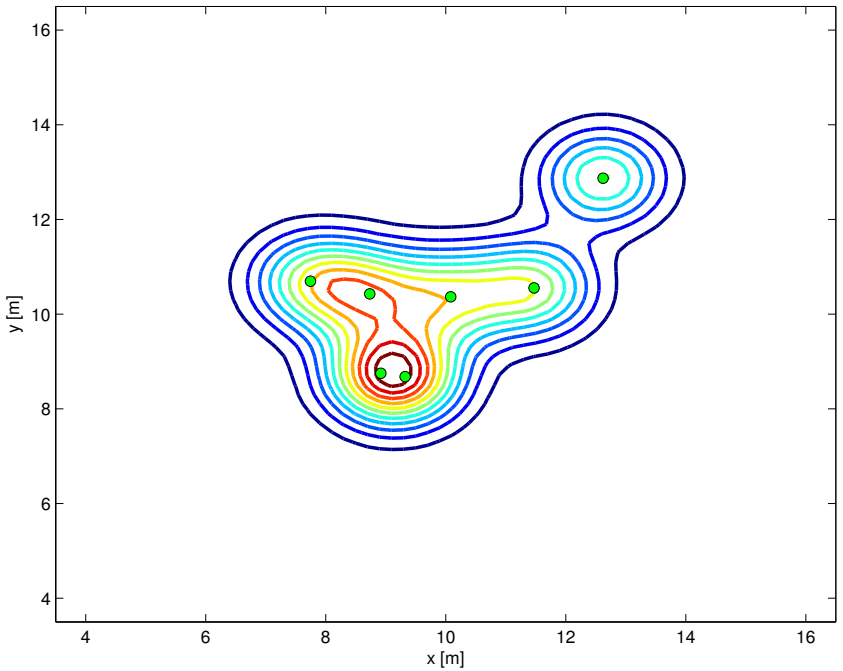
- For the extracted pedestrian estimates $\hat{x}_{k|k}^{(i)}$
- Two pedestrian estimates i and j are connected if
  1. $\left\| \hat{\mathrm{p}}_{k|k}^{(i)} - \hat{\mathrm{p}}_{k|k}^{(j)} \right\|_2 < 2$ m and
  2. $\left\| \hat{\mathrm{v}}_{k|k}^{(i)} - \hat{\mathrm{v}}_{k|k}^{(j)} \right\|_2 < 1$ m/s
- Two pedestrian estimates k and ℓ belong to the same group g if
  1. there is a set of $N \geq 2$ components indexed $i_1, i_2, \ldots, i_N$,
  2. where $i_1 = k$ and $i_N = \ell$,
  3. and where i_s and i_{s+1} are connected for $s = 1, \ldots, N-1$.
- The number of pedestrians in a group g is estimated as

$$\hat{N}_{k|k}^{g} = \sum_{j=1}^{J_{k|k}^{g}} w_{k|k}^{g,(j)}.$$

V Threshold PHD-intensity to represent shape

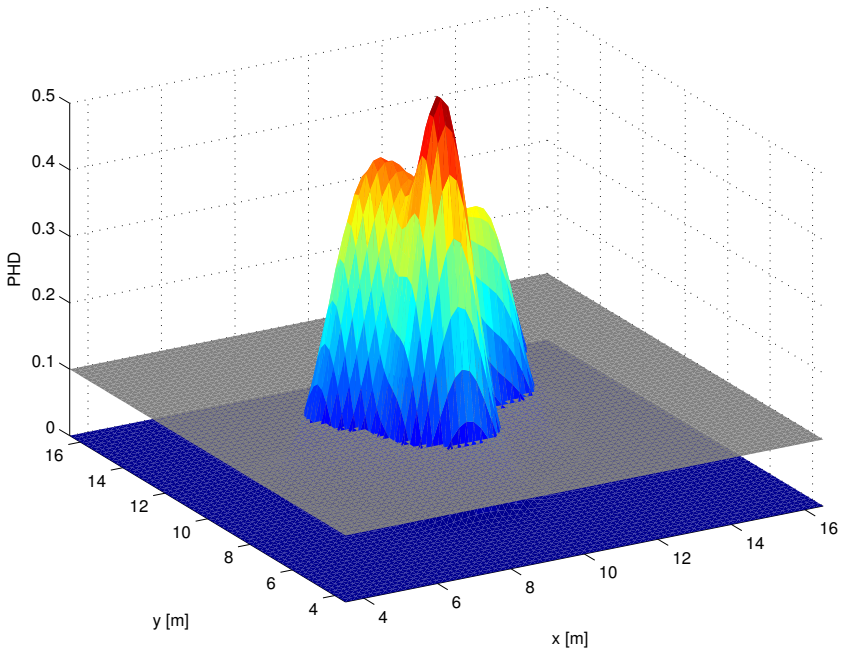
- For group g the PHD-intensity is $\sum_{i=1}^{J_{k|k}^g} w_{k|k}^{g,(i)} \mathcal{N}\left(x; m_{k|k}^{g,(i)}, P_{k|k}^{g,(i)}\right)$

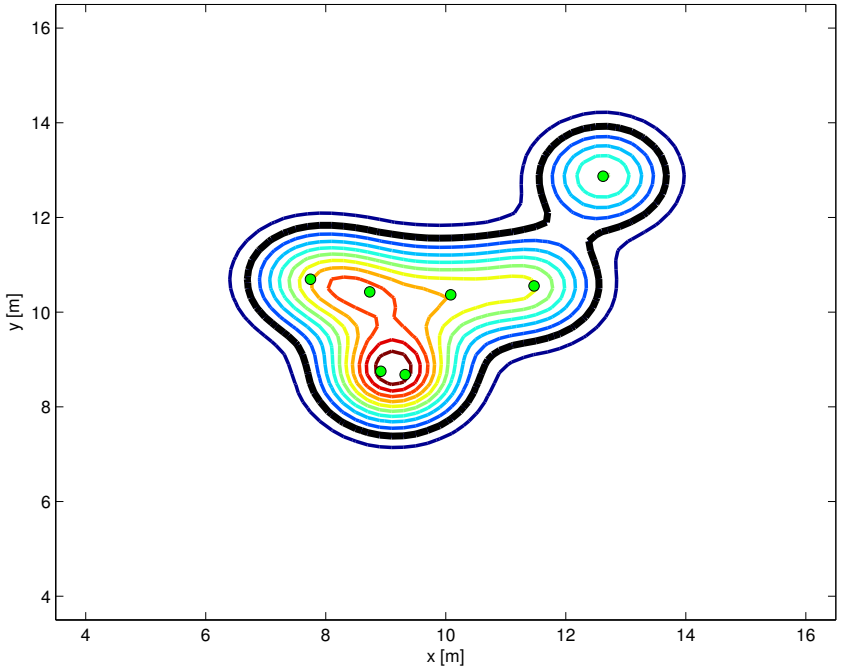




V Threshold PHD-intensity to represent shape

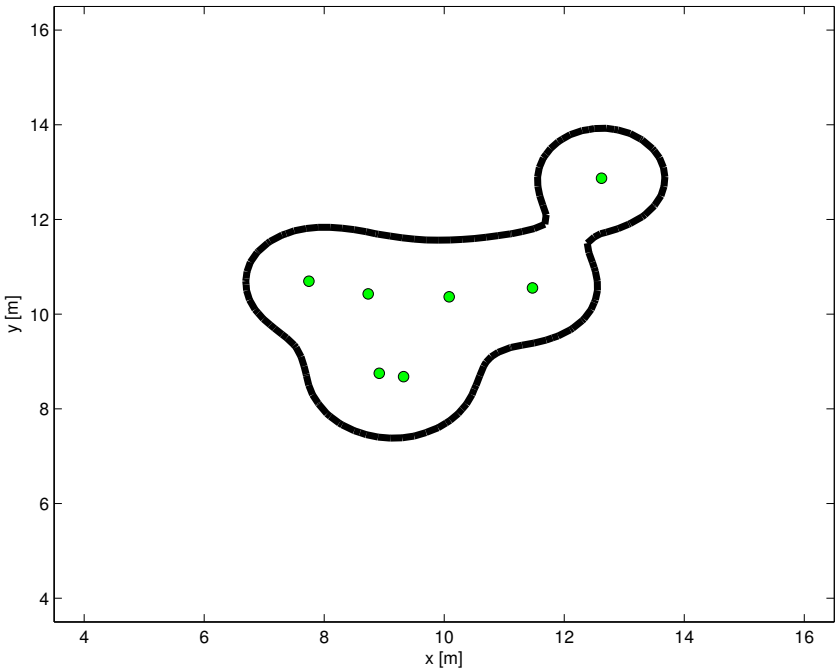
- For group g the PHD-intensity is $\sum_{i=1}^{J_{k|k}^g} w_{k|k}^{g,(i)} \mathcal{N}\left(x; m_{k|k}^{g,(i)}, P_{k|k}^{g,(i)}\right)$
- PHD is intersected at 0.1
  Intersection is approximation of group's size and shape.

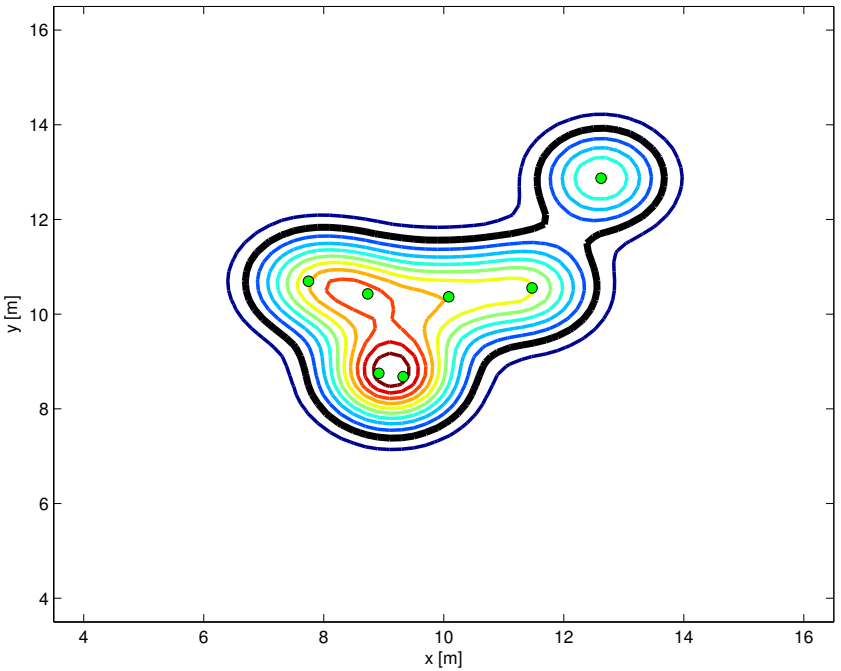




V Threshold PHD-intensity to represent shape

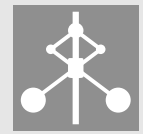
- For group g the PHD-intensity is $\sum_{i=1}^{J_{k|k}^g} w_{k|k}^{g,(i)} \mathcal{N}\left(x; m_{k|k}^{g,(i)}, P_{k|k}^{g,(i)}\right)$
- PHD is intersected at 0.1
  Intersection is approximation of group's size and shape.



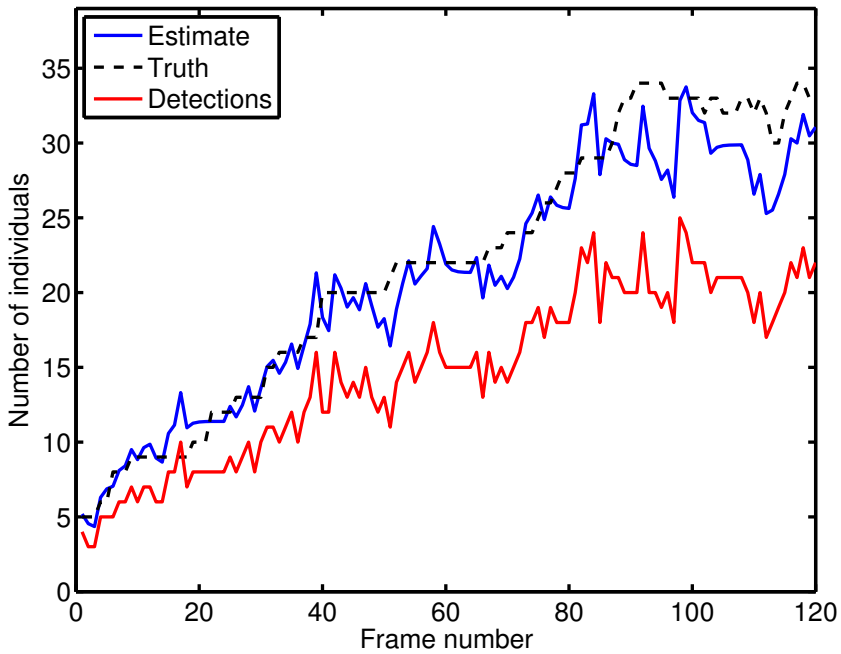


Results – video

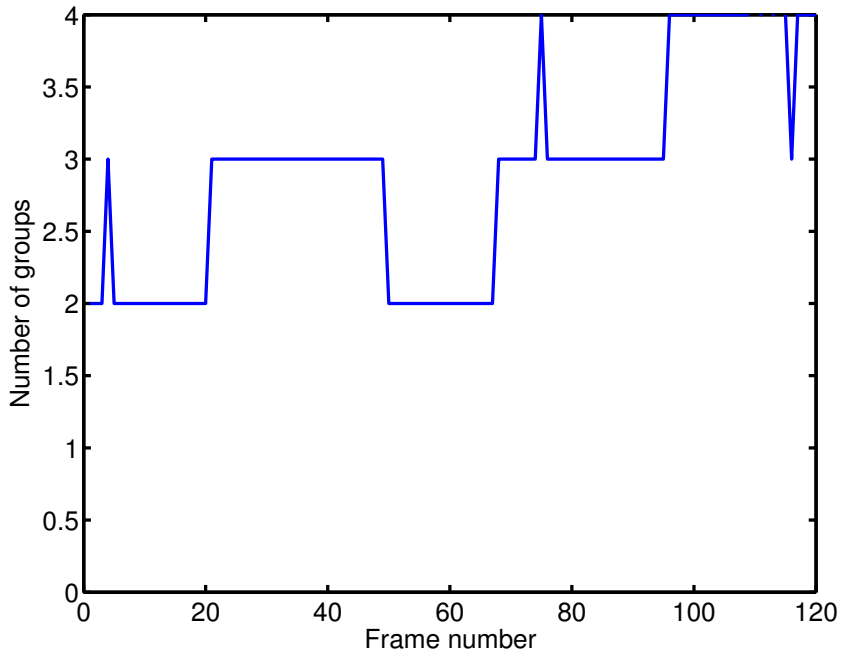
http://youtu.be/aAz3poW49CU



Results – number of individuals



Number of pedestrians



Number of groups

Summary

- Pedestrian detector used on each image
- Detections projected onto ground plane
- Projections input in GM-PHD filter to get pedestrian estimates
- Groups formed using clustering
- Group shape represented by thresholded PHD-intensity

Recent and future work

Recent work

- Gaussian mean shift clustering to form groups
- Group IDs for continuity
- See Viktor Edman's Master's Thesis

  `http://urn.kb.se/resolve?urn=urn:nbn:se:liu:diva-93996`

  `http://youtu.be/7DEkaTeFLWY`

Future work

- Compute uncertainty of image-to-ground plane projection
- Compare results with extended target PHD-filter
- Include also group dynamics, not just individual dynamics
- Dynamic way of forming groups

Thank you for listening!

Any questions?